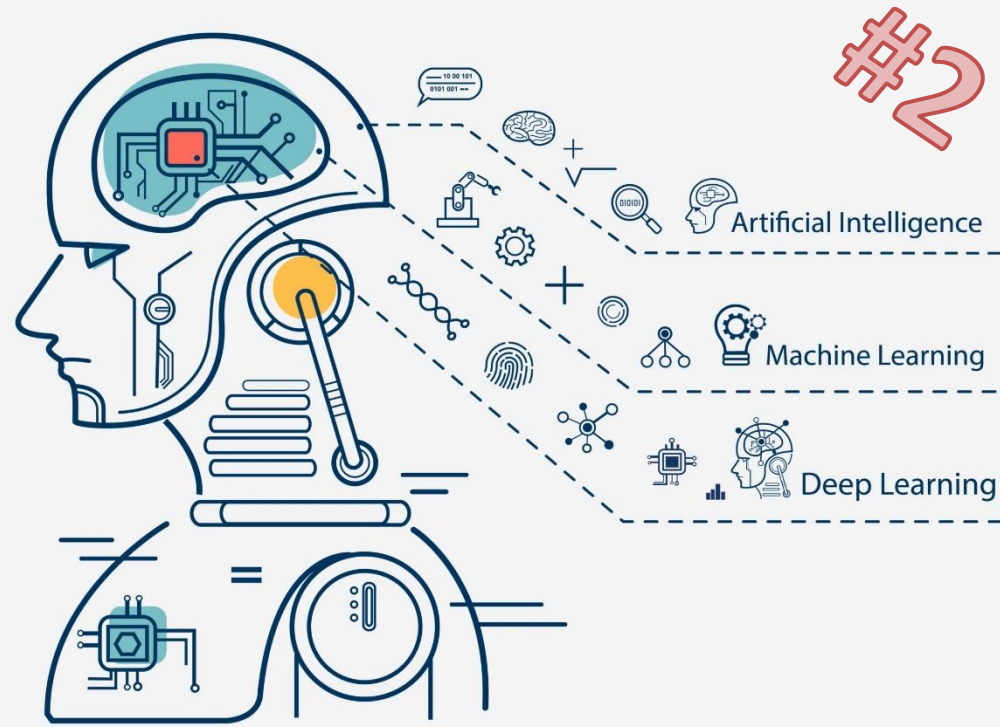




Machine Learning

Per Madsen

Faglærer

E-mail: pmad@techcollege.dk

Stor tak til Lærke Brandhøj Kristensen

Agenda

Slide 2

- Opfølgning på øvelser
 - #1 Indgang med Python
 - #2 Data cleaning and filtering
- Øvelse
 - #3 Visualisering og WordCloud
- Lidt teori om Machine Learning
- Øvelse (Start)
 - #4 Spam / Ham
 - #5 Spam / Ham / Målpinde

Hvordan lærer mennesker?

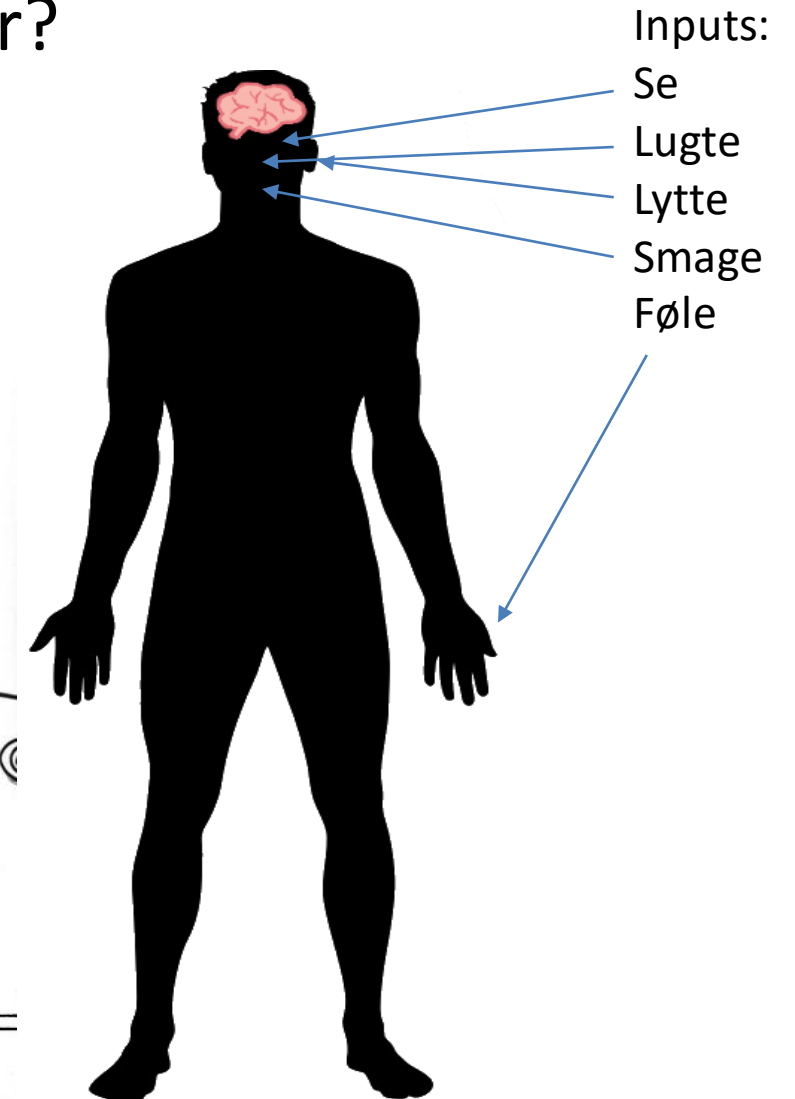
- Vi lærer baseret på vores sanser.
- Erfaring dannes ud fra oplevelser fra sanser.

Hvordan lærer vi af erfaring?

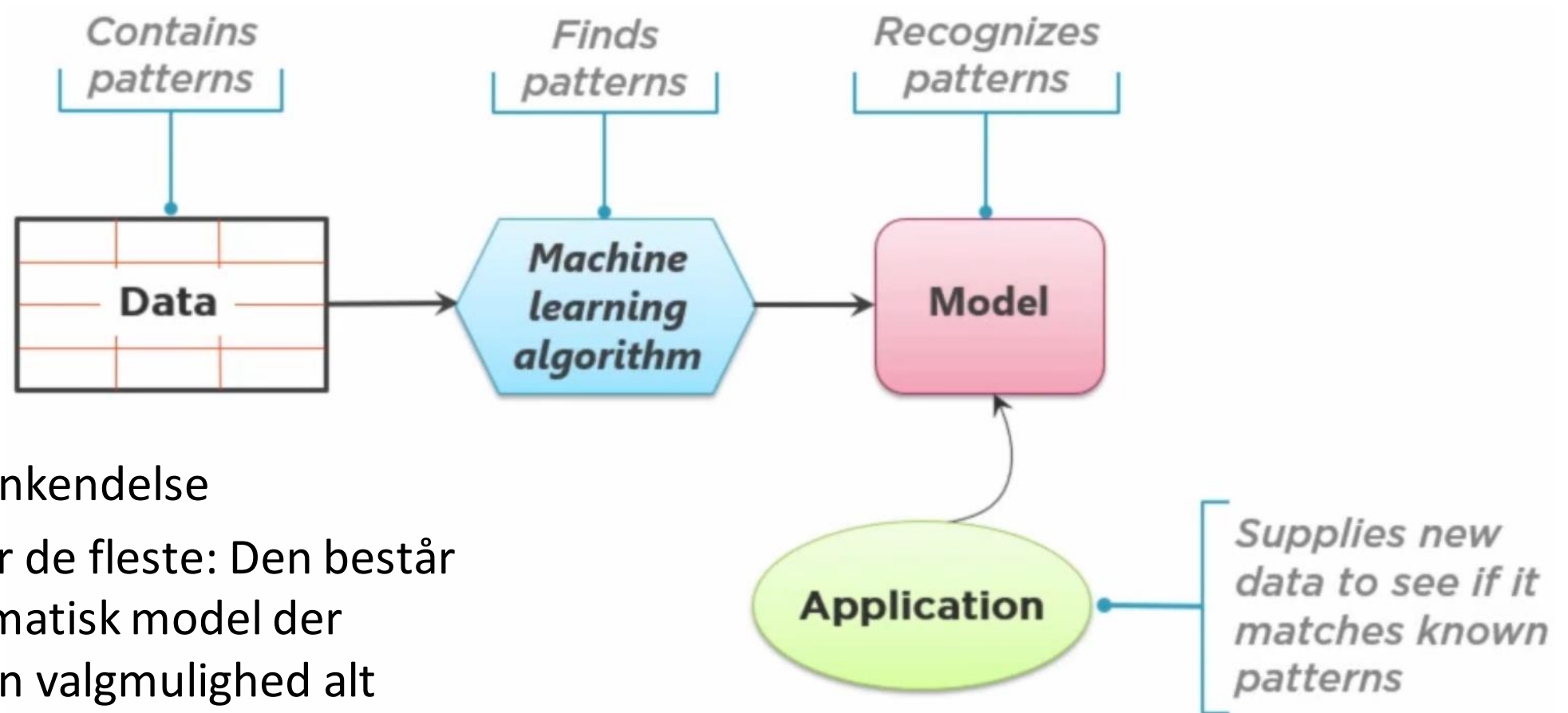
- Vi oplever gode og dårlige sanse inputs som former vores handlinger fremover.

Hvordan lærer vi at læse?

- Vi ser mønstre i tekst og lærer at genkende mening bag ord og sætninger.



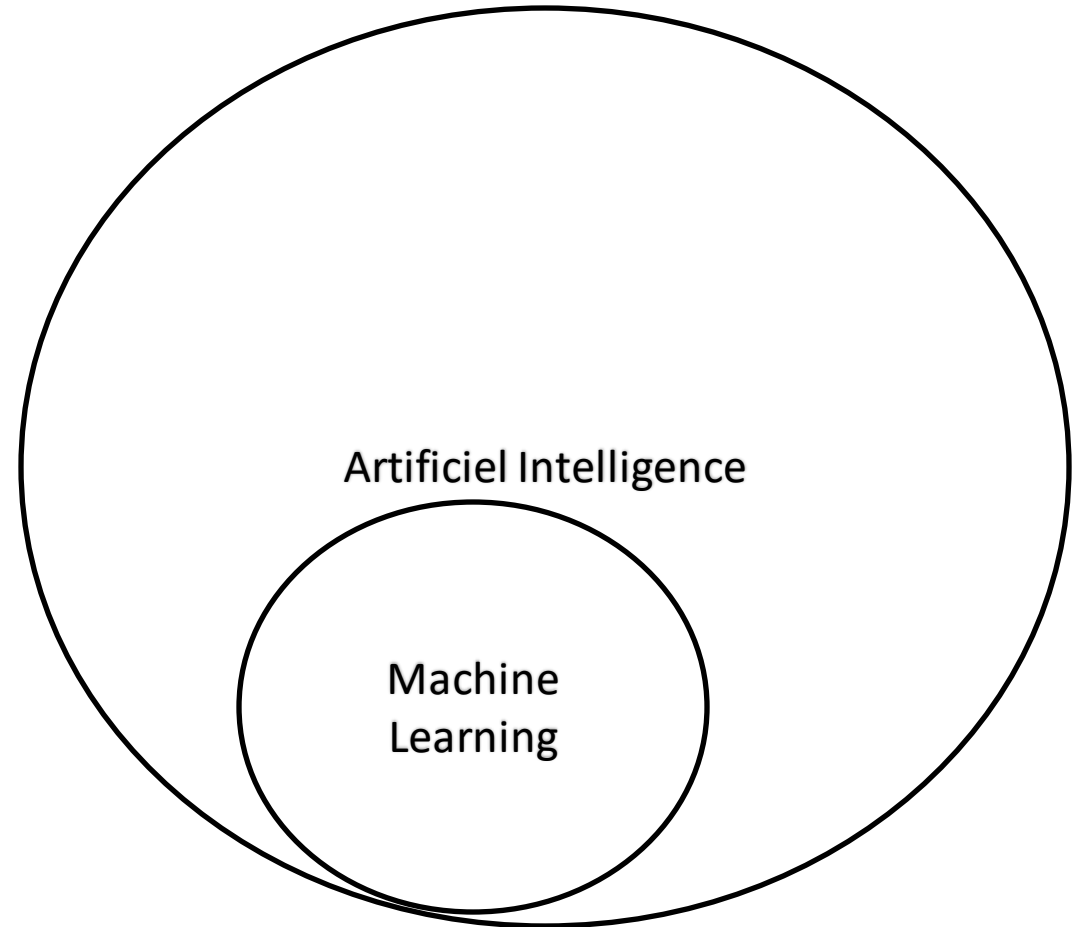
Hvad er Machine Learning?



- Mønster genkendelse
- Blackbox for de fleste: Den består af en matematisk model der outputter en valgmulighed alt efter input.

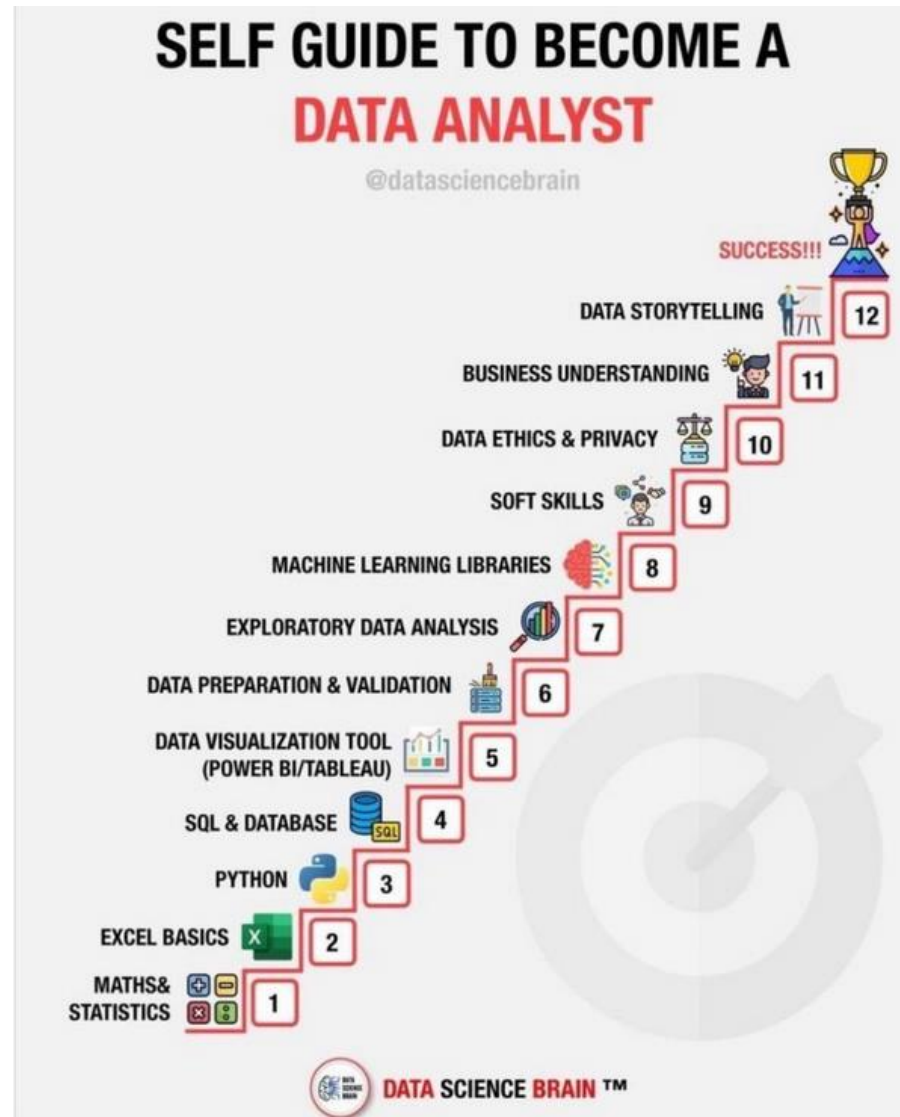
AI vs. ML

- Artificial Intelligence (AI)
 - Få maskiner til at "tænke"
 - Strong AI
 - Artificial General Intelligence (AGI)
 - Weak AI
 - Begrænset til specifik anvendelse
- Machine Learning (ML)
 - Lær fra data
- Verden er meget større end hvad dette fag kommer igennem (eller hvad underviseren ved)



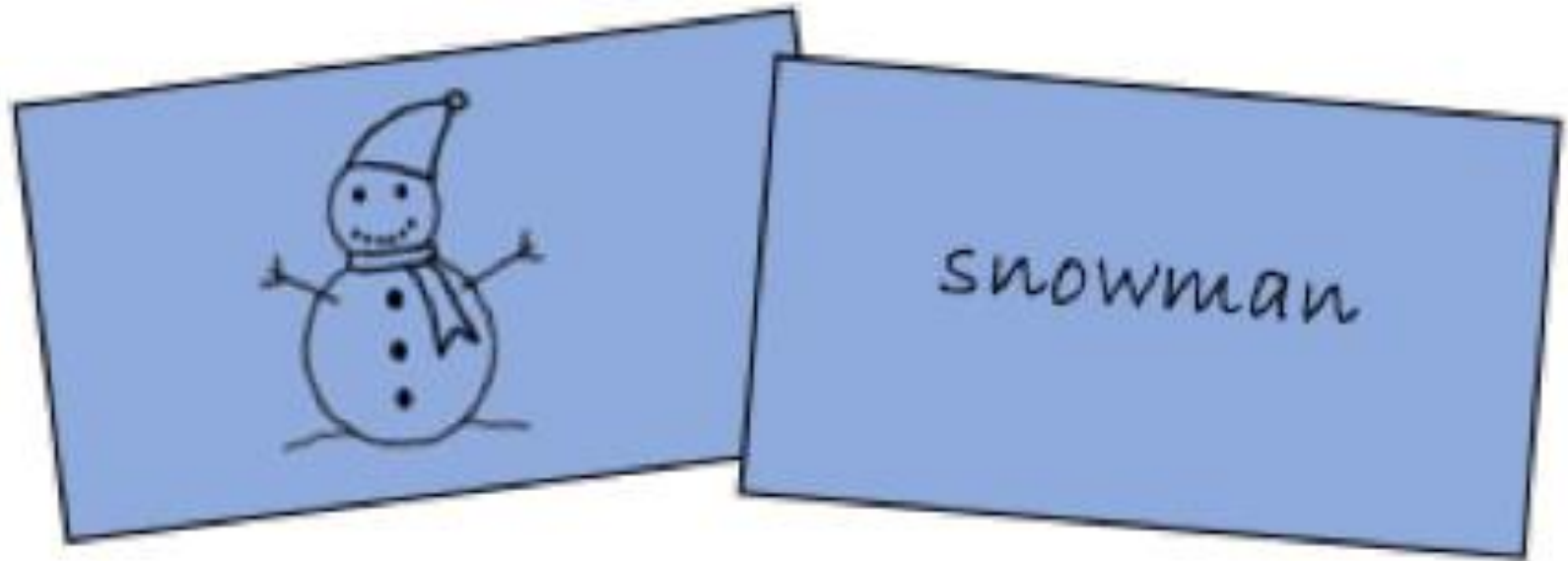
En del af en større rejse

Slide 6



4 paradimer: Supervised

- Supervised
 - Når man træner modellen, ved at vise den svaret.



Flash card

4 paradimer: Unsupervised

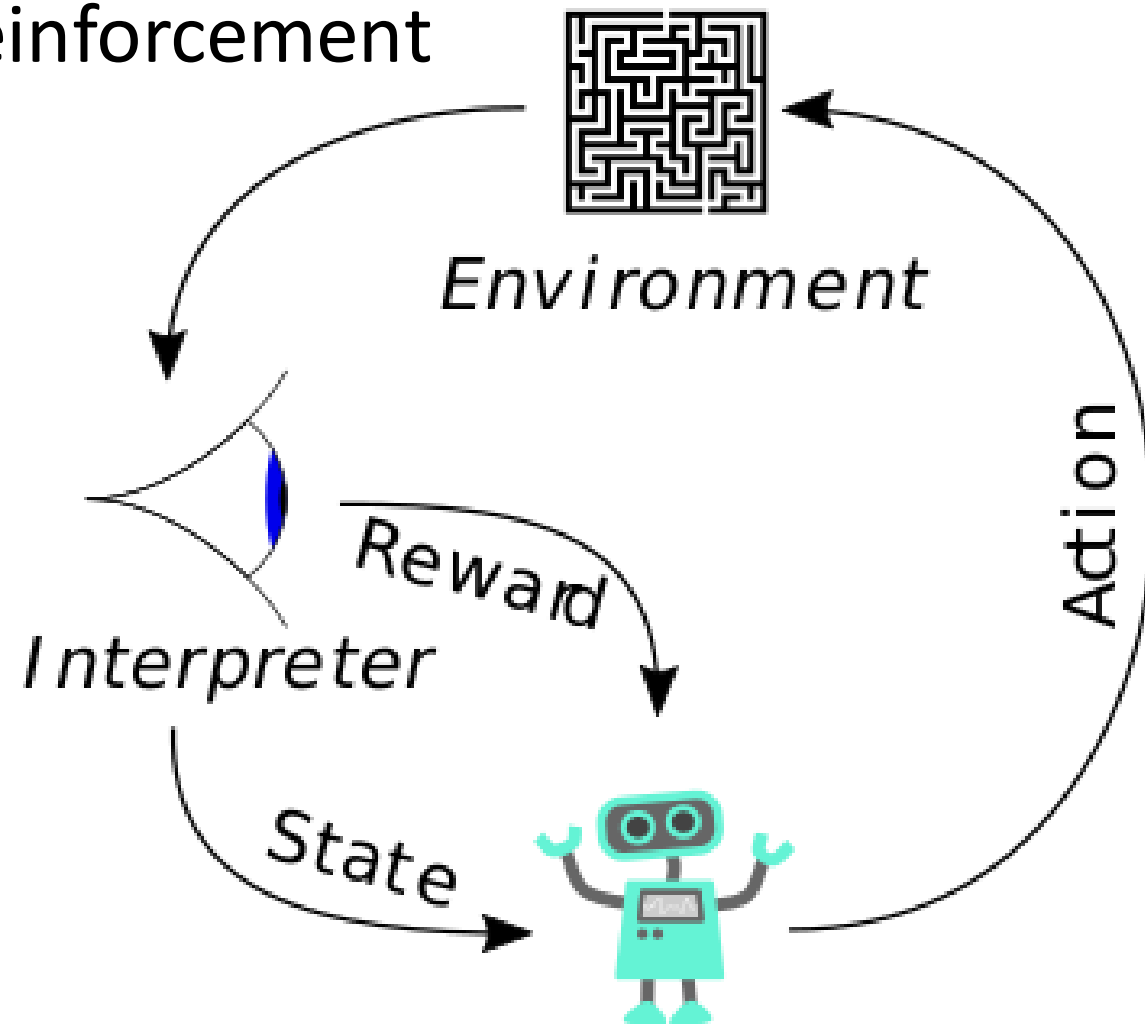
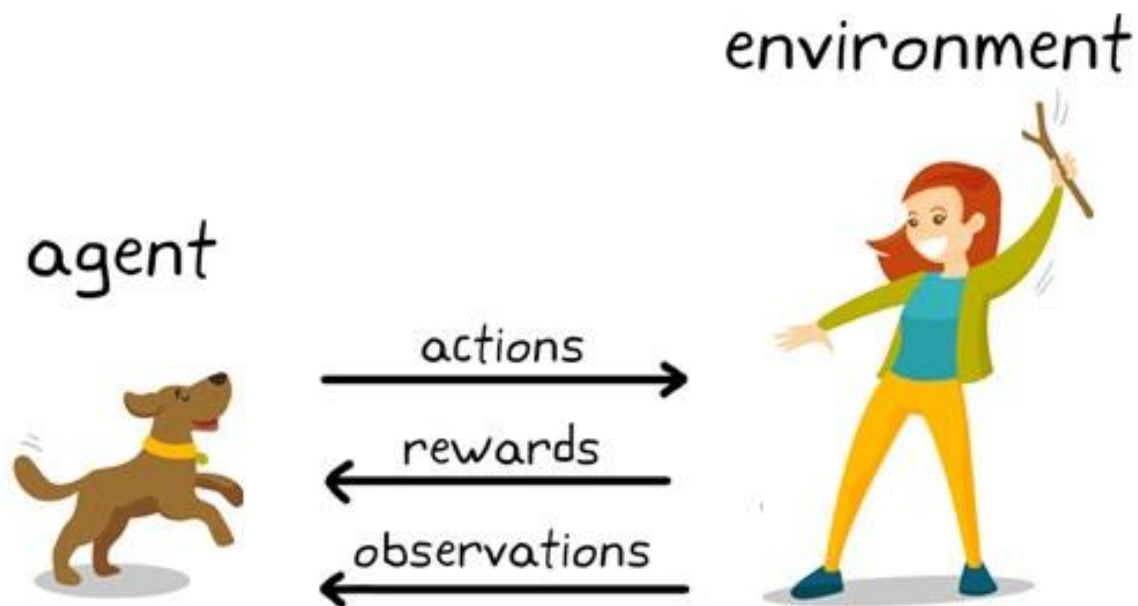
- Unsupervised
 - Når man træner modellen, hvor den ikke har svaret.

Repetition is the mother of all learning.
Repetition is the mother of all learning.
Repetition is the mother of all learning.
Repetition is the mother of all learning.
Repetition is the mother of all learning.
Repetition is the mother of all learning.
Repetition is the mother of all learning.

4 paradimer: Reinforcement

- Reinforcement

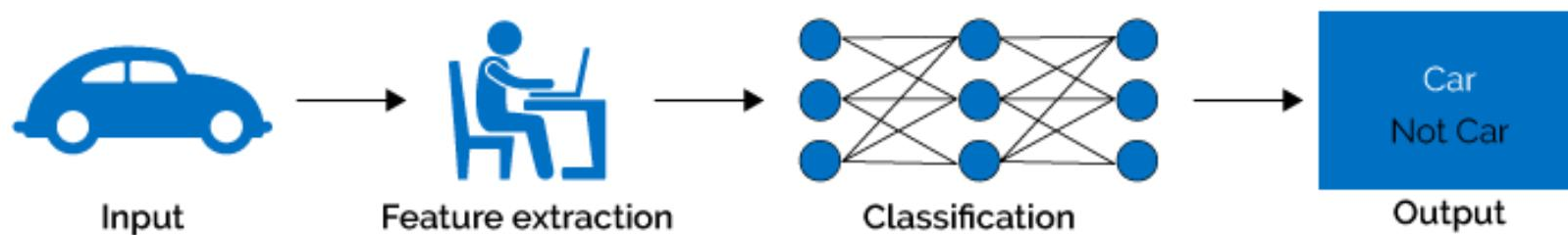
- Når man træner modellen, ved at sætte den i et miljø og giver den feedback baseret på dens handlinger



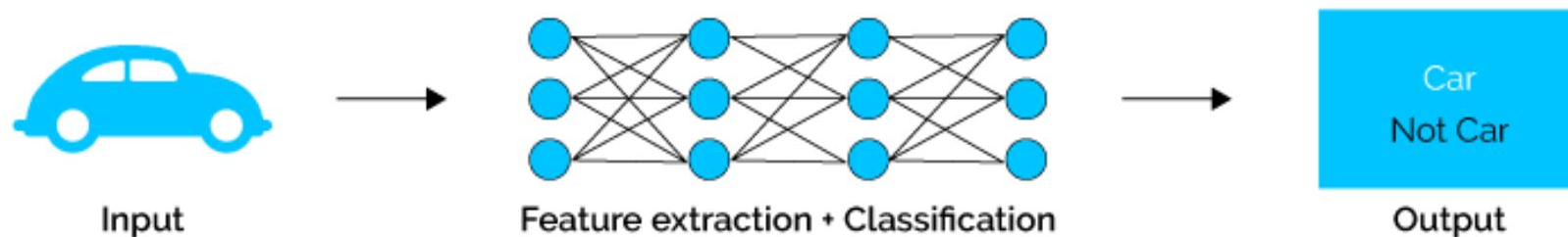
4 paradimer: Deep learning

- Deep learning algoritmer finder selv sammenhængen i dataen
- Det gør dem meget dårlige at bruge til at løse specifikke opgaver
- De er gode til at finde sammenhæng i kæmpe mængder af data, som vi ikke kan se en sammenhæng i.
- Neurale netværk: Layers, bias, weights, hidden layers

Machine Learning



Deep Learning



Learning by doing

- Øvelse #4: Spam / Ham
- Supervised Learning
- Tutorial med faldgruber
- Prøve at bruge **scikit-learn** uden at kende teorierne bag

Data processing

- Import the required packages
- Loading the Dataset
- Remove the unwanted data columns
- Preprocessing and Exploring the Dataset
- Build word cloud to see which message is spam and which is not.
- Remove the stop words and punctuations
- Convert the text data into vectors

Building a sms spam classification model

Slide 13

- Split the data into train and test sets
- Use Sklearn built-in classifiers to build the models
- Train the data on the model
- Evaluate model(s)
- Make predictions on new data



Adobe Stock | #72757953

GO!